



## INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

<b>(51) International Patent Classification <sup>6</sup> :</b> C07H 21/00, 21/04, C07K 2/00, 14/00, C12N 5/10, 5/16, 7/01, 15/00, 15/10, 15/11, 15/12, 15/63, C12P 1/00, 21/00, 21/02	<b>A1</b>	<b>(11) International Publication Number:</b> <b>WO 95/18140</b> <b>(43) International Publication Date:</b> 6 July 1995 (06.07.95)
<b>(21) International Application Number:</b> PCT/US94/14554 <b>(22) International Filing Date:</b> 28 December 1994 (28.12.94)  <b>(30) Priority Data:</b> 108205 28 December 1993 (28.12.93) IL  <b>(71) Applicant (for all designated States except US):</b> YEDA RE- SEARCH & DEVELOPMENT CO., LTD. [IL/IL]; Weiz- mann Institute of Science, P.O. Box 95, 76100 Rehovot (IL).  <b>(71) Applicant (for VN only):</b> RYCUS, Avigail [US/IL]; 16 Kipnis Street, 76305 Rehovot (IL).  <b>(72) Inventors; and</b> <b>(75) Inventors/Applicants (for US only):</b> BEN-ARIE, Nissim [IL/IL]; 105 Eshkol Street, 76804 Bazkeret Batya (IL). LANCET, Doron [IL/IL]; 4A Melchett Street, 65215 Tel Aviv (IL).  <b>(74) Agents:</b> COHEN, Herbert et al.; Wigman, Cohen, Leitner & Myers, P.C., Suite 200, Crystal Square 3, 1735 Jefferson Davis Highway, Arlington, VA 22202 (US).		<b>(81) Designated States:</b> CA, JP, US, VN, European patent (AT, BE, CH, DE, DK, ES, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE).  <b>Published</b> <i>With international search report.</i>
<b>(54) Title:</b> OLFACTORY GENES AND RECEPTORS		
<b>(57) Abstract</b>  Olfactory genes appear as gene clusters. In the human genome such clusters appear on chromosomes 11 and 17. Such gene clusters are disclosed herein for the first time. A variety of DNA sequences encoding olfactory receptors or parts thereof have been found and are disclosed herein.		

**FOR THE PURPOSES OF INFORMATION ONLY**

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AT	Austria	GB	United Kingdom	MR	Mauritania
AU	Australia	GE	Georgia	MW	Malawi
BB	Barbados	GN	Guinea	NE	Niger
BE	Belgium	GR	Greece	NL	Netherlands
BF	Burkina Faso	HU	Hungary	NO	Norway
BG	Bulgaria	IE	Ireland	NZ	New Zealand
BJ	Benin	IT	Italy	PL	Poland
BR	Brazil	JP	Japan	PT	Portugal
BY	Belarus	KE	Kenya	RO	Romania
CA	Canada	KG	Kyrgyzstan	RU	Russian Federation
CF	Central African Republic	KP	Democratic People's Republic of Korea	SD	Sudan
CG	Congo	KR	Republic of Korea	SE	Sweden
CH	Switzerland	KZ	Kazakhstan	SI	Slovenia
CI	Côte d'Ivoire	LI	Liechtenstein	SK	Slovakia
CM	Cameroon	LK	Sri Lanka	SN	Senegal
CN	China	LU	Luxembourg	TD	Chad
CS	Czechoslovakia	LV	Latvia	TG	Togo
CZ	Czech Republic	MC	Monaco	TJ	Tajikistan
DE	Germany	MD	Republic of Moldova	TT	Trinidad and Tobago
DK	Denmark	MG	Madagascar	UA	Ukraine
ES	Spain	ML	Mali	US	United States of America
FI	Finland	MN	Mongolia	UZ	Uzbekistan
FR	France			VN	Viet Nam
GA	Gabon				

## OLFACTORY GENES AND RECEPTORS

### FIELD OF THE INVENTION

The present invention is generally in the field of human genetics and concerns genes of the human olfactory system.

### 5 PRIOR ART

The following is a list of prior art referred to in the present  
specification:

1. Buck, L. and Axel, R., 1991, *Cell* 65:175-187.
- 10 2. PCT Application WO 92/17585.
3. Levy, N.S., *et al.*, 1991, *J. Steroid Biochem. Molec. Biol.*, 39:633-637.
4. Buck, L.B., 1992, *Current Opinion Genet. Devel.*, 2:467-473.
5. Parmentier, M., *et al.*, 1992, *Nature*, 355:453-455.
- 15 6. Nef, P., *et al.*, 1992, *Proc. Natl. Acad. Sci. USA*, 89:8948-8952.
7. Selbie, L.A., *et al.*, 1992, *Mol. Brain Res.*, 13:159-163.
8. Raming, K., *et al.*, 1993, *Nature*, 361:353-356.

9. Ressler, K.J., *et al.*, 1993, *Cell*, 73: 597-609.
10. Ngai, J., *et al.*, 1993, *Cell*, 72:657-666.
11. Schurmans, S., *et al.*, 1993, *Cytogenet Cell Genet*, 63:200-204.
- 5 12. Nizetic, D., *et al.*, 1991, *Proc. Natl. Acad. Sci. USA*, 88:3233-3237.
13. Kwiatkowski, D.J., *et al.*, 1990, *Am. J. Hum. Genet*, 46:559-567.
14. Pinkel, D., *et al.*, 1986, *Proc. Natl. Acad. Sci. USA*, 83:2934-2939.
15. Fidlerová, H., *et al.*, 1993, *Submitted*.
16. Ledbetter, D.H., *et al.*, 1989, *Proc. Natl. Acad. Sci. USA*, 86:5136-  
10 5140.
17. Rackwitz, H.R., *et al.*, 1985, *Gene*, 40:259-266.

The acknowledgement herein of the above publications is given to allow an appreciation of the prior art but should not, however, be  
15 construed as an indication that this art is in any way relevant to the patentability of the invention as defined in the appended claims.

The above publications will be acknowledged herein by indicating their number from the above list.

---

## 20 BACKGROUND OF THE INVENTION

The genetic basis for the olfactory system has been subject to extensive research. An understanding of the genetic basis of olfaction is important, for example, for the fragrance industry, to enable a more systematic and accurate way of fragrance design. To date, research in this  
25 field has been centered primarily in non-human mammals, although some information on the genetic basis of the olfactory system in humans has also been found.

On the basis of growing body of evidence, it became clear that the olfactory receptors (ORs) are encoded by a superfamily consisting of a

plurality of genes<sup>1-9</sup>. Common characteristics of all receptors identified to date is that they belong to the superfamily of G-protein coupled receptors and have seven putative trans-membrane helices.

Genomic blot analyses have shown that the rat genome may have several hundred OR genes. It is anticipated that large OR gene repertoires exist also in other mammals. In addition to the study of OR genes in rats<sup>1,3,8</sup>, members of the OR gene family have also been cloned from fish<sup>10</sup>, mouse<sup>6,9</sup>, dog<sup>5</sup> and human<sup>5,7,11</sup>.

It is an object of the present invention to provide novel DNA molecules encoding human olfactory receptors.

It is more specifically an object of the invention to provide a human olfactory subgenome, i.e. a cluster of OR genes.

It is another specific object of the present invention to provide novel proteins or polypeptides encoded by the DNA sequences of the invention.

The remaining objects of the invention will be realized in the following description and claims.

---

## GENERAL DESCRIPTION OF THE INVENTION

In accordance with the present invention it has been found that contrary to the prior art prevailing belief that known OR genes are randomly distributed throughout the genome, human OR genes form continuous clusters in the genome (hereinafter "*OR gene clusters*"). One such OR gene cluster was found on human chromosome 17. This OR gene cluster was found to be about 0.35 Mb long with an average inter-genic separation between the genes of about 15 kb.

Furthermore, in accordance with the present invention novel DNA coding sequences from OR genes belonging to the OR cluster on chromosome 17 have been identified and cloned. These sequences are

depicted in Fig. 4 which shows 16 sequences of which those designated 17-2 and 17-4 are known and the rest are novel.

The present invention thus provides, by a first of its aspects, a DNA molecule selected from the group consisting of:

- 5 (a) an isolated cluster of olfactory receptor genes;
  - (b) an isolated non-coding fragment of the DNA sequence of (a);
  - (c) a DNA sequence included in (a) or (b) and included within a sequence coding for an olfactory receptor with the proviso that the sequences designated 17-2 and 17-4 in Fig. 4 are excluded;
  - 10 (d) a DNA sequence different than the DNA sequence of (c) and encoding the same protein or polypeptide;
  - (e) a DNA sequence of (c) or (d) in which one or more coding triplets from the open reading frame have been added, deleted or replaced, the protein or polypeptide encoded by said sequence
  - 15 having similar properties to that encoded by the DNA sequence of (c) or (d); and
  - (f) a DNA sequence capable of hybridizing under conditions of stringency to the DNA sequence of (a), (b), (c), (d) or (e).
- 

20 A specific olfactory receptor gene cluster in accordance with the invention is that located on human chromosome 17 or that located on human chromosome 11.

By a further aspect of the invention there is provided a DNA molecule having a coding sequence being a member selected from the group consisting of:

- 25 (a) a DNA sequence depicted in Fig. 4 with the proviso that the sequences designated 17-2 and 17-4 be excluded;
- (b) a DNA sequence encoding a functional polypeptide being a fragment of the polypeptide encoded by the DNA sequence of (a);

- (c) a DNA sequence comprising the DNA sequence of (a) or (b);
- (d) a DNA sequence of (a), (b) or (c) wherein one or more triplets in the open reading frame have been added, replaced or deleted, the protein or polypeptide encoded by the sequence having essentially the same properties as the protein encoded by the respective sequence of (a), (b) or (c);
- (e) a DNA sequence of (a), (b), (c) and (d) comprising also non-coding stretches of nucleic acid residues; and
- (f) a DNA sequence capable of hybridization under conditions of stringency with the DNA sequence of (a), (b), (c) or (d).

The above DNA molecules can be used in transforming a host cell to express the protein or polypeptide encoded by said DNA molecule. For this purpose, the DNA molecule is incorporated into an expression vector which is introduced into suitable host cells, which may be prokaryotic or eukaryotic. Eukaryotic host cells are generally preferred since the post-translational modification in prokaryotic and eukaryotic host cells is different, and thus expression in a eukaryotic host cell will bring to an expression product which more closely resembles that encoded by the olfactory genomic DNA in humans. A particular example of a vector and a host cell is the Baculovirus vector and an insect host cell, respectively.

The present invention thus provides a method of production of a protein or a polypeptide in a host cell comprising transforming a host cell by a DNA molecule selected from those defined above, growing the host cells under conditions allowing for expression of the protein or polypeptide encoded by said DNA molecule and harvesting the expressed product.

In accordance with yet another aspect of the invention, there is provided a protein or polypeptide encoded by the above coding sequence and being a member selected from the group consisting of:

- (a) a protein or polypeptide having an amino acid sequence depicted in Fig. 3;
- (b) a protein or polypeptide wherein one or more amino acid residues has been added, deleted or chemically modified, the protein or polypeptide having essentially the same biological properties as the respective protein or polypeptide under (a); and
- (c) a protein comprising a polypeptide having a sequence of (a) or (b).

The present invention also provides, by a further of its aspects, uses of the above DNA molecules, proteins or polypeptides.

Another use of the DNA molecule is in the construction of probes for use in studying olfactory polymorphism in a population or of probes for Southern blot hybridization and restriction fragment length polymorphism for detecting interindividual differences in olfactory receptors or olfactory clusters in humans. Useful for this purpose are characterizing fragments of an OR gene cluster or DNA sequence which hybridize thereto. The integrity of gene clusters is usually preserved over generations, i.e. gene clusters are usually inherited as a single unit. This is so since the recombination between pairs of chromosomes has a much higher probability of occurring outside such clusters. In such clusters, unique non-coding sequences of a length of a few tens of nucleic acid basis can usually be found, which can be used to identify the entire cluster. By the use of such fragments or nucleic acid sequences which can hybridize thereto, the distribution of OR gene cluster in a population and hence OR polymorphism, can be assayed.

Probes to assay for OR polymorphism are very important in the fragrance industry since on the basis of knowledge of such polymorphism, fragrances for use by specific populations may be used. Based on the type of olfactory receptors, different people may be responsive to different fragrances. At times, a homogeneous population may have many common



ORs and thus fragrance preparations specific for this population may be produced. Alternatively, by the knowledge of polymorphism in a population, a fragrance preparation can be produced which will create a positive response in most individuals.

5           The proteins or polypeptides of the invention may be used in probing for the presence of a certain fragrance, aroma or flavor in a medium. Another use of said polypeptides or proteins is in the development of fragrances for human use. When used in probing or in the development of fragrances, said protein or polypeptide may be immobilized onto a solid  
10 support or may be embedded into a suitable membraneous system, e.g. membrane of a host cell transformed with a DNA encoding for said protein or polypeptide displaying same on its surface, a liposome wherein said protein or polypeptide is embedded, and the like. In such use, binding of a fragrance, aroma or a flavor to said protein or polypeptide will give rise  
15 to a detectable signal. The manner of use of a receptor protein so that binding of a ligand thereto will give rise to a detectable signal is generally known *per se* and the description of such manner is outside the scope of the present writing.

---

20           A further use of the above protein and polypeptide is in raising antibodies. Such antibodies, which also form an aspect of the invention, may be used to detect and label ORs *in vivo* as well as in *in vitro* preparations. For such use, the antibodies may carry a detectable marker. Another use of the antibodies is in assays for the detection of the presence of fragrances in a medium. In such use, antibodies which recognize the  
25 fragrance binding site are brought into contact with said proteins or polypeptides in the presence of a tested medium, the level of bound or free antibodies being directly proportional to the level of fragrance in the tested medium.

As will no doubt be appreciated by the artisan, the DNA molecules and proteins or polypeptides of the invention may have various other uses, all being within the scope of the present invention.

The invention will now be illustrated in the following description  
5 of specific embodiments with occasional reference to the annexed drawings.

### BRIEF DESCRIPTION OF THE DRAWINGS

Fig. 1 shows fluorescence *in situ* hybridization (FISH) of cosmid 7 to metaphase spreads of human chromosomes. The arrows in the figure  
10 indicate labelling at the telomeric end of chromosome 17 on the two sister chromatids of both chromosomes.

Fig. 2 shows the mapping of the OR cluster in human 17p13.3. Thirty-six representative OR-positive cosmids from the human chromosome 17 library were assembled into 3 contigs utilizing a partial digest mapping  
15 technique. The *Hind*III (H) and *Eco*RI (E) sites are shown on the top lines. The numbered cosmids are shown as horizontal line, showing their positions in the contig. Gap lengths between the 3 contigs were estimated to be 10-20 kb by FISH to free chromatin, and are shown to scale. OR coding  
20 regions mapped to within  $\pm 4$ kb are shown as dark boxes, while those for which only a cosmid assignment is known are shown as open boxes. The approximate location of STS 506 (D17S126) and the PCR marker HP5 (in the 3' non coding region of OR17-OR17-210) are also shown. None of the cosmids tested positively with the flanking chromosome 17 STS's PFN1 and 508<sup>13</sup>. The approximate DNA breakpoints of Miller-Dieker Syndrome  
25 patient BR8 is shown, within cosmid 7 by Southern blot analysis of DNA from a cell line containing chromosome 17 of this patient on mouse background, with radiolabelled cosmid 7 DNA;

Fig. 3 shows the deduced amino acid sequences of olfactory receptor clones from human chromosome 17 and their alignment. The deduced

protein sequences of the PCR-generated clones encoding putative human ORs were aligned using the GCG package pileup program (Genetic Computer Group, Madison, Wisconsin, U.S.A.). Rat OR I14<sup>1</sup> is also included for comparison and positioning of the sequences in the full length open reading frame. The numbering system used is based on the I14 sequence. Transmembrane domains were determined by an averaged hydropathy plot procedure combined with a hydrophobic moment analysis and are marked by lines with roman numerals. Dark boxes represent amino acids identity, while different shades of gray stand for two different degrees of conservative substitutions. Marked by a • are positions of full identity among all the sequences. ORs 17-23 and 17-90 have a deletion of one nucleotide compared to all other sequences, which leads to an immediate stop codon to be in frame. Therefore, amino acid 166 in these sequences was marked as \*, and the remainder of the protein is the translation of the phase deduced by homology to the other OR sequences;

Fig. 4 shows the DNA sequences of the olfactory receptors from the human chromosome 17 cluster shown in Fig. 3; and

Fig. 5 shows the expression of OR17-93 in human olfactory epithelium. PCR analysis was performed using OR17-93 specific primers on the following templates; OR17-93 and OR17-2 clones as positive controls (OR), cDNA of olfactory epithelium (OR), cDNA of respiratory epithelium (RE), cDNA from nasal inverted papilloma (PA). Arrows are pointing at the PCR products of OR17-93 and OR17-2 clones (260 and 230 bp, respectively). M=100 bp ladder markers (BRL, Gaithersburg, MD).

## DESCRIPTION OF A SPECIFIC EMBODIMENT

The invention will now be described by outlining the experiments which led to the discoveries which form the basis of the present invention.

This section is intended for illustration of the invention and should not be construed to limit the invention in any way.

## Materials and Methods

### 5 OR Probe preparation

Genomic DNA was extracted from male Wistar rats and from blood of a human individual. The rat probe (ORI) was obtained by PCR using oligonucleotides OR5B (TM2; CCC ATG TA(T/C) TT(G/C/T) TT(C/T) CTC (A/G/T)(G/C) (C/T) AA(C/T)(T/C)T(G/A) TC C) and OR3B  
10 (TM7; OR3B AG(A/G) C(A/T)(A/G) TAI ATG AAI GG(A/G) TTC AIC AT). The human probe, ORIII, was amplified by nested PCR, using first OR5B/OR3B and then reamplified by OR5A (GGC (C/T)TA TGA CCG IT(A/T) (T/C)(G/C)T (G/A)GC (C/T)AT ITG) and OR3B. PCR amplification was carried out in a buffer containing 50mM KCl, 10mM Tris pH 8.3,  
15 0.01% gelatine, 0.2 mM each deoxyribonucleotide, 1 $\mu$ M each primer and 2.5 U of Taq DNA polymerase (Cetus, Branchburg, NJ) 100 $\mu$ l as follows: 30 cycles of denaturation at 94°C, annealing at 55°C and extension at 72°C, each step for 1 min. The first step of denaturation and the last step of extension were 4 min. long. Both PCR-generated probes were purified by  
20 GeneClean II (BIO 101, La Jolla, CA).

### Library Screening

A 5-fold coverage Chromosome 17 cosmid library was used, prepared previously from flow-sorted chromosome 17 DNA, isolated from  
25 human cell line LCL127<sup>12</sup>. DNA probes were radiolabelled to a specificity activity of 10<sup>8</sup>-10<sup>9</sup> cpm/ $\mu$ g by random hexamer priming using [ $\alpha$ -32P] dCTP (Amersham, Buckinghamshire, UK). Hybridization to filters was carried out overnight in 50% formamide, 50 mM sodium phosphate buffer, pH 7.2, 1 mM EDTA, 10X Denhardt's, 4X SSC, 1% SDS, 10% dextran

5 sulphate, 50  $\mu\text{g/ml}$  sonicated salmon sperm DNA, and 100  $\mu\text{g/ml}$  sonicated human placental DNA at 42°C. Cosmid and genomic filters were washed in 1X SSC, 0.1% SDS at room temperature for 60 minutes and exposed to Kodak XAR 5 X-ray film at -80°C. The positive cosmids isolated were numbered sequentially from 1 to 76. The isolation of 70 OR-positive cosmids (average insert length 35 kb) in a region of 350 kb suggests a 7-fold coverage of the library, over this region, in agreement with the previously estimated coverage.

#### 10 'Partial digest' mapping of the cosmids

Cosmid DNA was purified using a Qiagen column (Diagen, Chatsworth, CA) according to the manufacturer's recommendations. For the cosmid blots, 1  $\mu\text{g}$  of DNA was digested by the appropriate restriction enzyme and 400 ng run on an 0.65% agarose gel. The DNA was visualized by staining with ethidium bromide and transferred onto nylon membranes (Hybond N+, Amersham, Buckinghamshire, UK), according to the manufacturer's recommendations. For partial digest mapping, 10  $\mu\text{g}$  of cosmid DNA was linearised by cleavage with lambda terminase (1U/mg DNA) for 2 hours at room temperature and 2  $\mu\text{g}$  partially digested with *EcoRI* or *HindIII* in a total volume of 30  $\mu\text{l}$  at 37°C (0.2 U/ $\mu\text{g}$  DNA, removing 5  $\mu\text{l}$  aliquots every 5 minutes). Partially digested cosmid DNA were hybridized in solution with radiolabelled oligonucleotides complementary to the protruding 12bp single-stranded sequences at the cleaved cos site and the samples fractionated on a 0.5% agarose gel. The annealing of labelled oligonucleotides specific for the 'left' or 'right' termini of the cos sites allowed the autoradiographic visualization of restriction fragments of increasing length extending from each side of the cosmid. A restriction map was constructed by calculating and compiling the digitized migration distances of partial digestion restriction fragments.

### Minimal Complexity Index (MCI) Analysis

In order to find how many OR coding regions lie in each cosmid, PCR amplification using OR5B and OR3B as primers was performed on individual cosmids. The PCR products were digested with a frequent cutter restriction enzymes (e.g. *Hinf*I, *Hae*III, *Hha*I). After agarose gel separation and ethidium bromide staining, the lengths of the fragments were summed and divided by the length of the undigested product, as described <sup>1,2</sup>. The value obtained is the Minimal Complexity Index (MCI), indicating the minimal number of putative receptors encoded by the cosmid.

### Sequencing of cloned PCR products

Cosmid DNA was extracted by the miniprep alkaline lysis procedure. PCR was performed separately on each, using OR5B/OR3B primers. An equimolar pool of PCR products from all cosmids, as well as products amplified from individual cosmids were ligated to pCR1000 or pCRII plasmids (Invitrogene, San Diego, CA). Recombinant clones were identified by Southern blot hybridization and PCR. Sequencing was performed using Taq DNA polymerase and dye-terminators on an ABI DNA sequencer model 373A, with universal vector primers and HOR1, an internal primer (GA/(C/T) G(A/G)(A/C) T(A/T)(C/T) (G/C)TG GCC ATG TGC).

### Fluorescence *in situ* hybridization (FISH)

Hybridization and detection were carried out according to a modification of the technique described by Pinkel *et al.*<sup>14</sup>. For metaphase *in situ* hybridization, 1  $\mu$ g of cosmid DNA was biotinylated using a Bionick kit (BRL, Gaithersburg, MD). 80 ng of labelled DNA was mixed with 4  $\mu$ g Cot-1 DNA (BRL, Gaithersburg, MD), dried and resuspended in hybridization buffer (50% formamide, 2xSSC, 10% dextran sulphate, 1% Tween 20).

The sample was denatured by heating to 80°C for 3.5 minutes and incubated at 37°C for 60 minutes. Chromosome slides were baked for two hours at 65°C and then denatured in 70% formamide, 2xSSC for 3 minutes at 75°C. Slides were dehydrated by washing with 70%, 95% and absolute ethanol.

5 Hybridization took place overnight at 37°C under sealed coverslip. Post hybridization washes were with 2xSSC, 50% formamide (3 x 5 min.) and 2xSSC (3 x 5 min.) both at 42°C. The slides were then washed in 4xSSC, 0.05% Tween 20, pH 7 (SSCT) for 3 minutes at 37°C and preincubated in SSCT containing 5% low fat dried milk (Marvel) (SSCTM) prior to the  
10 addition of FITC-conjugated avidin DCS (Vector labs) at 5 µg/ml in SSCT. The signal was amplified by addition of biotinylated anti-avidin (5 µg/ml) and a second application of FITC-avidin (DCS (5 µg/ml). Finally the slides were washed with phosphate buffered saline and dehydrated through an ethanol series.

15 Chromatin was released from routinely fixed nuclei according to a recently developed method<sup>15</sup>. For two color FISH one probe was labelled with biotin and one with digoxigenin, as above. Detection was by incubation with mouse anti-digoxigenin monoclonal antibody (Boehringer  
Manheim, Manheim, Germany) at 0.8 µg/ml plus avidin Texas red at  
20 2 µg/ml, both in SSCTM. This was followed by anti-mouse IgG conjugated to FITC at 30 µg/ml. The biotinated anti-avidin was used alone and finally another round of avidin Texas red was performed. Post hybridization washes were as described. Slides were mounted in Citiflour (Citiflour Ltd., London) with propidium iodide 0.5 µg/ml for one color visualization or  
25 DAPI (diamidino-2-phenyl-indole dihydrochloride; Sigma, St. Louis, MO) 0.1 µg/ml for two colors, and photographed using a Zeiss Axiophot microscope.

### Somatic cell hybrid mapping

Somatic cell hybrids BR8 and AY1 were constructed by fusion of lymphoblastoid cell lines of Miller-Dieker syndrome patients MDS-9 and MDS-12, respectively, with TK<sup>+</sup> rodent cells followed by selection in HAT media. Hybrid clones containing a single, detected chromosome 17 were identified by RLFP analysis of polymorphic loci in distal 17p and confirmed by cytogenetic analysis.

### OR tissue expression

Human tissues were surgically removed from a female patient having inverted papilloma and were immediately frozen in liquid nitrogen. Total RNA was extracted using TRI-REAGENT (NRC Inc., Cincinnati, OH) following the manufacturer's procedure. Yields were 3.3, 0.9 and 1.6 mg RNA/g tissue for papilloma, olfactory epithelium and respiratory epithelium, respectively. Total RNA (2  $\mu$ g) was treated with 0.2 units reverse transcriptase (BRL, Gaithersburg, MD) for 20 mins. at 37°C and was amplified using OR17-93 specific primers (OR17-93A: ATT GCT GCA GAA CAT GCA GAA; OR17-93B: ACA TGG CAT GGA AGG TGG TC).

A control experiment was conducted to verify that the PCR signals are RNA-dependent, and not due to a genomic DNA contamination. Since Taq DNA polymerase is known to have a significant reverse transcriptase activity, the omission of reverse transcription was considered less desirable. Instead, control PCR reactions were performed using primers for  $\beta$ -actin spanning intron 2.

### Results

#### OR-positive cosmids from human chromosome 17

Conserved regions in the 2nd and 7th transmembrane domains of the published rat OR sequences<sup>1,2</sup> were used to design oligonucleotide



primers (OR5B and OR3B, respectively). These were utilized to amplify a complex set of OR-related sequences by polymerase chain reaction (PCR), using rat and human genomic DNA as templates (probes ORI and ORIII, respectively). The resulting PCR products were radiolabelled and used to

5 screen cosmid libraries prepared from flow-sorted human chromosomes 17, 21, 22 and X<sup>12</sup>. A large number of positively hybridizing clones was obtained only for chromosome 17: 75 cosmids out of approximately 20,000 tested gave a clear hybridization signal with the human genomic probe (ORIII), of which about half (36) contained OR genes sufficiently conserved

10 to be detected by the ORI rat PCR probe as well. The clones were individually tested for the presence of OR coding sequences by Southern hybridization and a PCR amplification with the primers described above. The latter is feasible since OR genes have no introns in their coding regions<sup>1,2,6</sup>. Seventy hybridization positive cosmids gave the expected ~ 700

15 bp long product generated by the primers OR5B and OR3B, and were further characterized. Each of the PCR products was then digested with frequent-cutting restriction enzymes, and subjected to a Minimal Complexity Index (MCI) analysis (see Materials and Methods), giving an estimate for

---

the minimal number of OR coding regions per cosmid. The results showed

20 an average of 2.3 ORs per ~ 35 kb cosmid genomic insert, or approximately an average of one OR coding region per 15 kb of genomic DNA. There was a correlation between the intensity of hybridization signal in the library screen and the inferred number of ORs present on a given cosmid (not shown).

25 Six cosmids showing disparate OR restriction patterns (numbers 1,2,4,7,8,10) were used for fluorescence *in situ* hybridization to metaphase spreads of human chromosomes (Fig. 1). All four mapped to a similar position, 17p13, close to the p-telomere of chromosome 17. This result, together with the frequent presence of more than one OR coding region per

cosmid provided initial indications that at least some of the identified OR genes were clustered.

### Chromosomal mapping of the OR gene cluster

5           The approximate size of the genomic OR locus was determined by hybridizing the 5B/3A PCR products from cosmids, 1,2,4 and 7 to a Southern blot containing human genomic DNA digested with a variety of infrequently cutting restriction enzymes (*BssHII*, *SacII*, *MluI*, *NruI* and *NotI*). All four PCR products identified a ~ 400 kb *BssHII* restriction  
10 fragment, giving an estimate for the maximal size of the region. This indicated that the clustered OR genes do not have CpG islands at their 5' flanking ends. *HindIII* and *EcoRI* restriction maps were constructed for 40 cosmids using a partial digest mapping protocol<sup>17</sup>, and 35 of them arranged into three contigs with a total length of 347 kb (Fig. 2). All the other  
15 cosmids were judged by full restriction digest patterns with *HindIII* to be overlapping with the mapped subset. Two small gaps were seen, which probably arise due to the large distance between neighboring OR genes in these regions. The gap sizes between cosmids 53-45, 53-68 and 28-65 were estimated at 10-20 kb, and the contigs oriented relative to each other,  
20 by two-color FISH<sup>15</sup> to free chromatin using pairwise combination of cosmids from the ends of the three established contigs (data not shown). This produced clear linear signals from cosmids hybridized to chromatin fibers, enabling the relationship between adjacent or overlapping probes to be determined.

25           A more accurate localization and an absolute orientation of the OR cluster within the chromosomal band 17p13.3 was obtained using DNA from somatic cell hybrids containing the deleted human chromosome 17 homolog of two Miller-Dieker Syndrome (MDS) patients<sup>16</sup>. DNA from hybrid BR8 was positive with the OR17-93 PCR primers but negative with

PCR marker HP5 (near OR17-210) as well as with STS marker 506<sup>16</sup> which is found only on cosmid 73 (Fig. 2). Southern hybridization with cosmids 7, 44 and 59 further confirmed the location of the BR8 breakpoint at the centromeric end of the OR cluster, within cosmid 7. DNA from hybrid  
5 AY1 was positive with all three PCR markers, consistent with its breakpoint being telomeric to STS marker 506. Thus, the OR cluster is localized between the BR8 breakpoint and STS 506, immediately centromeric to the MDS critical region.

#### 10 Mapping and sequencing of OR coding regions

In order to localize the OR genes in the cluster, and to obtain DNA sequence information for them, an equimolar mixture of coding regions amplified by PCR (with primers OR5B-OR3B) from each of the 70 cosmids, was used to generate a library of OR clones. Of 17 clones  
15 sequenced, 11 were different. Restriction patterns of OR5B-OR3B PCR products with up to six frequent-cutting restriction enzymes allowed a comparison of these clones to those seen in PCR products from the individual cosmids. This suggested the existence of 6 additional OR genes  
in the cluster, and provided cosmid assignments for each of the 16 OR  
20 coding regions identified. By relating this information to the cosmid physical map, it was possible to localize most of the receptors with an accuracy of  $\pm 4$  kb within the cluster. Further positional information was obtained from Southern blot hybridization of the ORIII probe to *HindIII*, and *HindIII/EcoRI* digests of the cosmids (Fig. 2).

25 The 17 OR protein sequences (between transmembrane domains 2 and 7) are shown in Fig. 3 and the corresponding DNA sequences are shown in Fig. 4. The ORs have 36-39% mutual amino acid sequence identity. Two of the sequences (OR17-2 and OR17-4) are identical to recently published human OR clones HUMHGM071<sup>5</sup> and HGM07E<sup>11</sup>,

respectively. The latter gene was independently mapped to 17p13-->p12<sup>11</sup>. All the OR gene sequences were co-linear, except OR17-93 which had a 9 amino acid insert between positions 129 and 130, within intracellular loop 2. In addition, there is a single amino acid deletion at position 268 in  
5 four of the OR sequences. In two cases, pairs of OR genes shared >99% identity, with members of each pair mapping to the same position within the accuracy of our determination. For the first pair (OR17-2 and OR17-31), the difference was one silent mutation (at amino acid 218). For the second pair (OR17-23 and OR17-90, both pseudogenes with the same single base  
10 deletion at amino acid 166) the difference was a trinucleotide deletion in OR17-90, which affected two coded amino acids (positions 243-4). These two OR gene pairs may be alleles, occurring at respectively identical loci on cosmids derived from the parental and maternal chromosomes. Alternatively, they could be the result of a recent tandem duplication.

15

### Tissue expression

The OR genes described here were isolated from cloned genomic DNA fragments. To examine whether any of them are actually expressed  
in human olfactory epithelium, samples of olfactory and non-sensory nasal  
20 tissues from a human subject were obtained. Total RNA was extracted, and examined for OR expression by a reverse transcriptase - polymerase chain reaction (RT-PCR) assay (Fig. 5). A control PCR with  $\beta$ -actin primers spanning an intron was used to demonstrate that signals are not significantly generated from a genomic DNA contamination. Since OR genes are rather  
25 similar to each other, it was necessary to verify that a specific OR of the presently described cluster is the one actually amplified. The verification made use of the unusual 27 nucleotide insertion in OR17-93 and this, together with the design of a specific primer pair allowed to ascertain that only one gene product was amplified.

## CLAIMS:

1. A DNA molecule selected from the group consisting of:
  - (a) an isolated cluster of olfactory receptor genes;
  - 5 (b) an isolated non-coding fragment of the DNA sequence of (a);
  - (c) a DNA sequence included in (a) or (b) and included within a sequence coding for an olfactory receptor with the proviso that the sequences designated 17-2 and 17-4 in Fig. 4 are excluded;
  - 10 (d) a DNA sequence different than the DNA sequence of (c) and encoding the same protein or polypeptide;
  - (e) a DNA sequence of (c) or (d) in which one or more coding triplets from the open reading frame have been added, deleted or replaced, the protein or polypeptide encoded by said sequence having similar properties to that encoded by the DNA sequence of (c) or (d); and
  - 15 (f) a DNA sequence capable of hybridizing under conditions of stringency to the DNA sequence of (a), (b), (c), (d) or (e).
2. A DNA molecule according to Claim 1, wherein the isolated gene cluster is derived from human chromosome 11 or 17.
- 20 3. A DNA molecule having a coding sequence being a member selected from the group consisting of:
  - (a) a DNA sequence depicted in Fig. 4 with the proviso that the sequences designated 17-2 and 17-4 be excluded;
  - 25 (b) a DNA sequence encoding a functional polypeptide being a fragment of the polypeptide encoded by the DNA sequence of (a);
  - (c) a DNA sequence comprising the DNA sequence of (a) or (b);
  - (d) a DNA sequence of (a), (b) or (c) wherein one or more triplets in the open reading frame have been added, replaced or deleted,
  - 30 the protein or polypeptide encoded by the sequence having

essentially the same properties as the protein encoded by the respective sequence of (a), (b) or (c);

- (e) a DNA sequence of (a), (b), (c) and (d) comprising also non-coding stretches of nucleic acid residues; and
- 5 (f) a DNA sequence capable of hybridization under conditions of stringency with the DNA sequence of (a), (b), (c) or (d).

4. A method of production of a protein or a polypeptide in a host cell comprising transforming a host cell by a DNA molecule of Claim 3, growing the host cells under conditions allowing for expression of the protein or polypeptide encoded by said DNA molecule and harvesting the  
10 expressed product.

5. A protein or polypeptide being a member selected from the group consisting of:

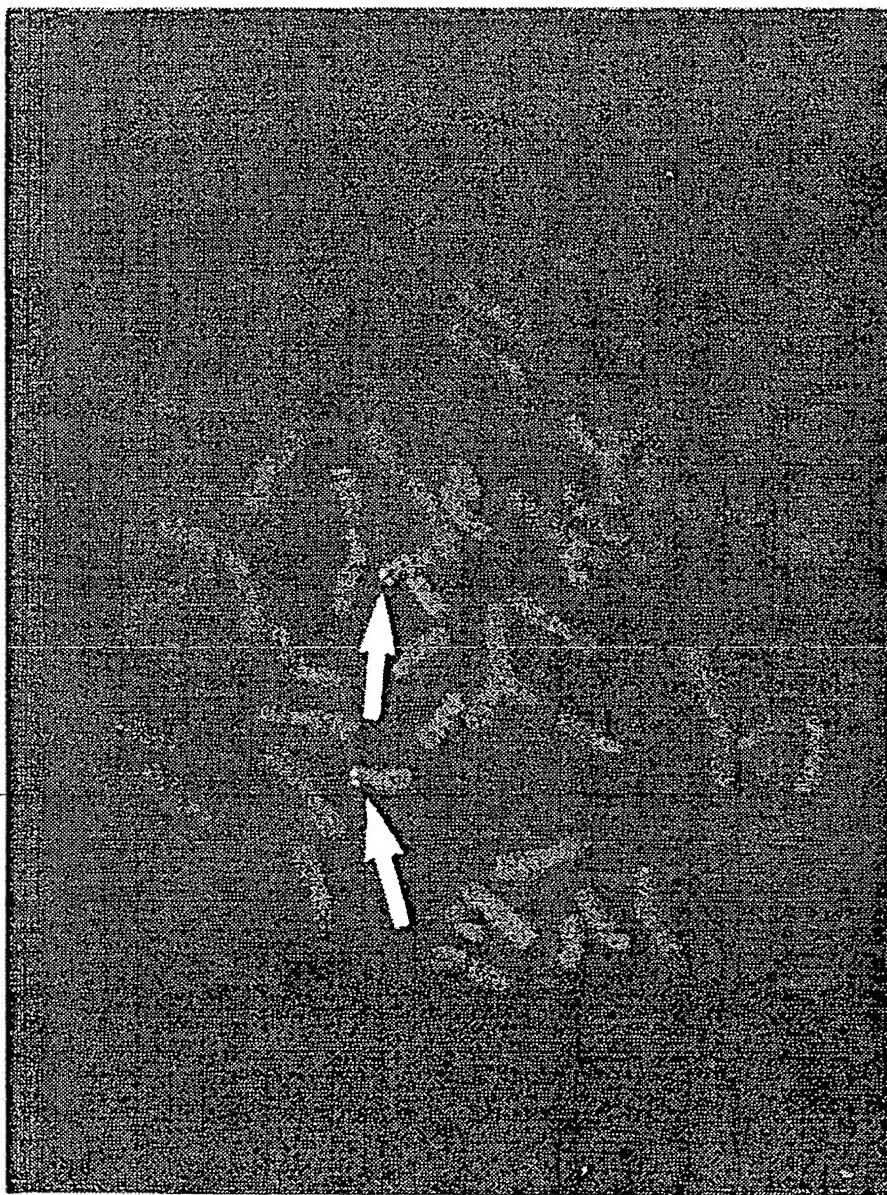
- 15 (a) a protein or polypeptide having an amino acid sequence depicted in Fig. 3;
- (b) a protein or polypeptide wherein one or more amino acid residues has been added, deleted or chemically modified, the protein or polypeptide having essentially the same biological properties as the respective protein or polypeptide under (a); and
- 20 (c) a protein comprising a polypeptide having a sequence of (a) or (b).

6. A host cell transformed by a DNA molecule according to any one of Claims 1 or 3.

7. A probe for use in assaying fragrances, aromas or flavors in a  
25 medium, comprising a protein or polypeptide according to Claim 5.

8. A probe for use in testing of the presence of a gene cluster among individuals in a population, comprising a DNA molecule as defined in Claim 1(b).

9. An antibody directed against a protein or polypeptide according  
30 to Claim 5.



**FIG. 1**

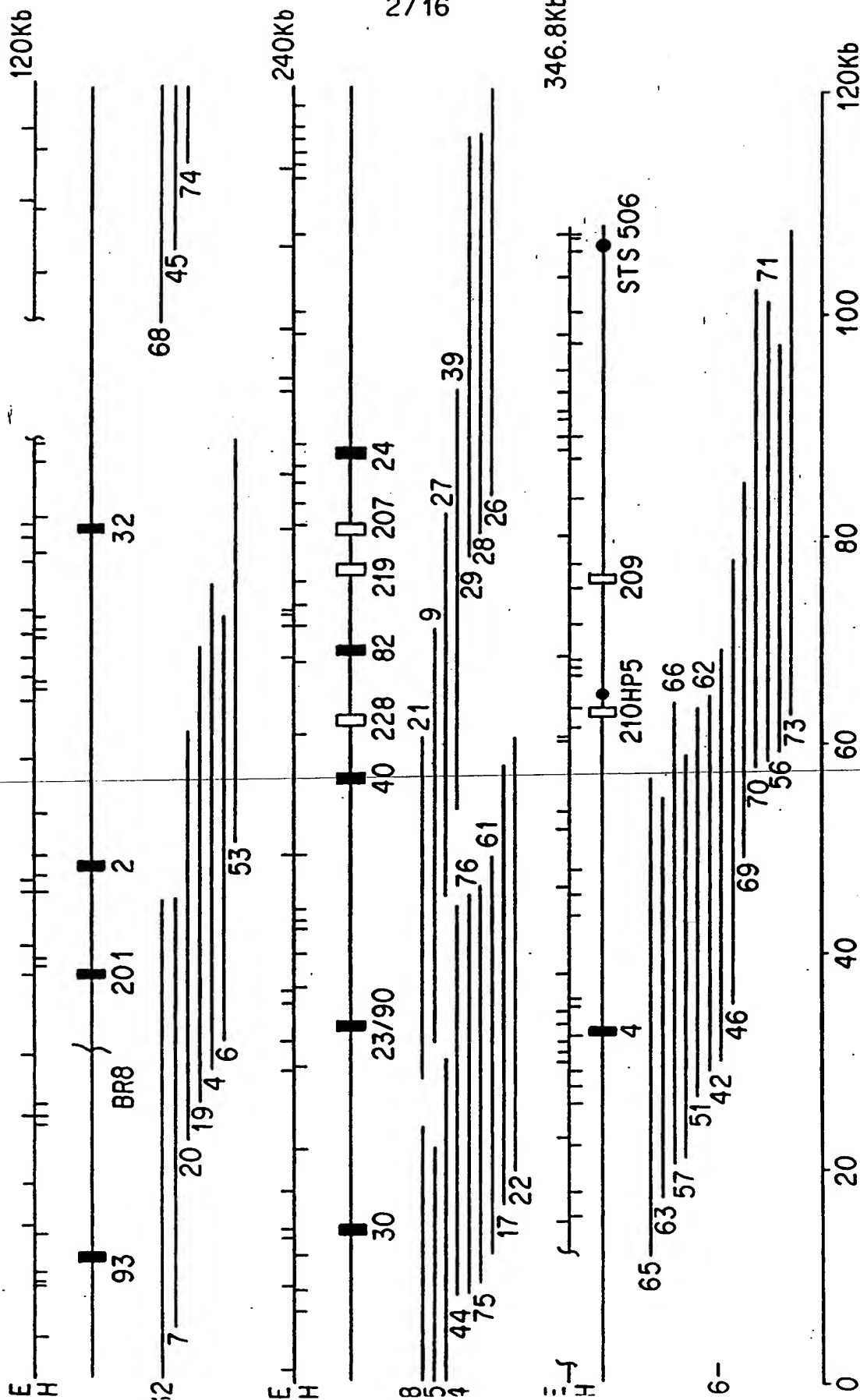


FIG. 2



II										III																																											
70										100										110																																	
80																																																					
114	F	S	D	L	C	F	S	S	V	T	M	P	K	L	L	Q	N	M	Q	S	Q	I	S	N	D	R	S	I	S	Y	T	G	C	L	T	Q	L	Y	F	F	M	V	F	G	D	M	E	S	F	I	L		
17-24	L	P	D	V	G	C	V	S	V	T	V	P	A	M	L	S	H	F	I	S	N	D	R	S	I	S	N	D	R	S	I	P	Y	K	A	C	L	S	E	L	F	H	L	L	A	G	A	D	C	C	F	I	L
17-219	L	P	D	V	G	C	V	S	V	T	V	P	A	M	L	S	H	F	I	S	N	D	R	S	I	S	N	D	R	S	I	P	Y	K	A	C	L	S	E	L	F	H	L	L	A	G	A	D	C	C	F	I	L
17-40	V	L	D	V	G	C	I	S	V	T	V	P	S	M	L	S	R	L	L	S	R	K	R	A	V	P	C	G	A	C	L	T	Q	L	Y	F	F	H	L	F	V	G	V	D	C	C	F	I	L				
17-82	V	L	D	V	G	C	I	S	V	T	V	P	S	M	L	S	R	L	L	S	R	K	R	A	V	P	C	G	A	C	L	T	Q	L	Y	F	F	H	L	F	V	G	V	D	C	C	F	I	L				
17-228	V	L	D	V	G	C	I	T	V	T	V	P	A	M	L	G	R	L	L	S	H	K	S	T	I	S	Y	D	A	C	L	S	Q	L	Y	F	F	H	L	L	A	G	M	D	C	C	F	I	L				
17-207	V	L	D	V	G	C	I	T	V	T	V	P	A	M	L	G	R	L	L	S	H	K	S	T	I	S	Y	D	A	C	L	S	Q	L	Y	F	F	H	L	L	A	G	M	D	C	C	F	I	L				
17-201	V	L	D	V	G	C	I	T	V	T	V	P	A	M	L	G	R	L	L	S	H	K	S	T	I	S	Y	D	A	C	L	S	Q	L	Y	F	F	H	L	L	A	G	M	D	C	C	F	I	L				
17-23	F	T	D	L	F	F	V	T	N	T	I	P	K	M	L	V	N	L	L	Q	S	Q	N	K	A	I	S	Y	T	G	C	L	T	Q	L	Y	F	L	V	S	L	V	A	L	D	N	L	N	L	N	L		
17-90	F	T	D	L	F	F	V	T	N	T	I	P	K	M	L	V	N	L	L	Q	S	Q	N	K	A	I	S	Y	T	G	C	L	T	Q	L	Y	F	L	V	S	L	V	A	L	D	N	L	N	L	N	L		
17-4	F	T	D	L	F	F	V	T	N	T	I	P	K	M	L	V	N	L	L	Q	S	Q	N	K	A	I	S	Y	T	G	C	L	T	Q	L	Y	F	L	V	S	L	V	A	L	D	N	L	N	L	N	L		
17-30	F	T	D	L	F	F	V	T	N	T	I	P	K	M	L	V	N	L	L	Q	S	Q	N	K	A	I	S	Y	T	G	C	L	T	Q	L	Y	F	L	V	S	L	V	A	L	D	N	L	N	L	N	L		
17-2	F	S	D	L	C	F	S	S	V	T	I	P	K	L	L	Q	N	M	Q	N	Q	D	P	S	I	P	Y	A	D	C	L	T	Q	M	Y	F	F	L	L	F	G	D	L	E	S	F	I	L					
17-32	F	S	D	L	C	F	S	S	V	T	I	P	K	L	L	Q	N	M	Q	N	Q	D	P	S	I	P	Y	A	D	C	L	T	Q	M	Y	F	F	L	L	F	G	D	L	E	S	F	I	L					
17-93	F	S	D	L	C	F	S	S	V	T	M	P	K	L	L	Q	N	M	Q	N	Q	D	P	S	I	P	Y	A	D	C	L	T	Q	M	Y	F	F	L	L	F	G	D	L	E	S	F	I	L					
17-210	F	S	D	L	C	F	S	S	V	T	M	P	K	L	L	Q	N	M	Q	N	Q	D	P	S	I	P	Y	A	D	C	L	T	Q	M	Y	F	F	L	L	F	G	D	L	E	S	F	I	L					
17-209	L	A	D	A	C	F	V	S	T	T	V	P	K	M	L	A	N	I	Q	I	Q	S	Q	A	I	S	Y	L	G	C	L	A	Q	M	Y	F	H	L	M	L	F	Y	G	V	L	E	S	F	I	L			
CONSENSUS	F	-	D	L	-	F	-	S	V	T	-	P	K	M	L	-	N	-	Q	S	Q	-	-	-	I	S	Y	-	-	C	L	T	Q	L	Y	F	-	-	L	-	-	L	D	-	-	L	D	-	-	F	I	L	

RECTIFIED SHEET

FIG. 3a

[illegible]

FIG. 3b



V		VI	
114	GGLI	210	250
17-24	A A F M	240	V V T L
17-219	G F I M	230	T T V V
17-40	G F I M	220	C C L L
17-82	G F I M	210	A A L L
17-228	G F I M	200	C C L L
17-207	G F I M	190	A A L L
17-201	A A F M	180	C C L L
17-23	G C F F	170	A A L L
17-90	G C F F	160	C C L L
17-4	G C F F	150	A A L L
17-30	G C F F	140	C C L L
17-2	G G L I	130	A A L L
17-32	G G L I	120	C C L L
17-93	G G L I	110	A A L L
17-210	G G L I	100	C C L L
17-209	G G L I	90	A A L L
ONSENSUS	G - - I	80	C C L L

FIG. 3d

7/16

		VII																				280
																						270
114		F	Y	G	T	I	F	G	I	Y	L	C	P	S	G	N	N	S	T	V	K	E
17-24		F	Y	G	T	G	V	F	F	Y	T	R	L	G	S	V	E	S	S	D	K	D
17-219		F	F	G	R	G	I	F	F	Y	M	R	L	G	S	E	E	A	S	D	K	D
17-40		F	Y	G	S	G	I	F	F	Y	M	R	L	G	S	E	E	L	S	D	K	D
17-82		F	F	G	R	G	I	F	F	Y	M	R	L	G	S	E	E	A	S	D	K	D
17-228		F	F	G	R	G	I	F	F	Y	M	R	L	G	S	E	E	A	S	D	K	D
17-207		F	Y	G	S	G	I	F	F	Y	M	R	L	G	S	E	E	L	S	D	K	D
17-201		F	Y	G	T	G	V	F	F	Y	M	R	L	G	S	V	E	S	S	M	K	D
17-23		F	Y	G	T	L	G	M	V	Y	L	Q	P	L	Q	T	Y	S	S	M	K	D
17-90		F	Y	G	T	L	G	M	V	Y	L	Q	P	L	Q	T	Y	S	S	M	K	D
17-4		F	Y	G	T	L	C	M	V	Y	L	Q	P	L	H	T	Y	S	S	M	K	D
17-30		F	Y	G	T	L	A	M	V	Y	L	Q	P	L	H	T	Y	S	S	M	K	D
17-2		F	Y	G	T	V	I	G	L	Y	L	C	S	S	A	N	S	S	T	L	K	D
17-32		F	Y	G	T	V	I	G	L	Y	L	C	S	S	A	N	S	S	T	L	K	D
17-93		F	Y	G	T	V	I	G	L	Y	L	C	P	S	A	N	S	S	T	L	K	D
17-210		F	Y	G	T	I	I	G	L	Y	L	C	P	L	T	N	H	N	T	V	K	D
17-209		F	F	G	T	S	F	C	V	D	F	S	S	P	S	T	H	S	A	Q	K	D
CONSENSUS		F	Y	G	T	I	I	I	I	Y	I	I	I	I	I	I	I	S	S	I	K	D

FIG. 3e

17-24

		10		20		30		40		50		60		70		80
1	CTGCCAGATG	TGGGTGTGT	CAGTGTCACT	GTCCTCGCCA	TGCTGAGCCA	TTTCATATCC	AACGACAGAA	GCATTCCCCTA	80							
81	TAAGGCCTGC	CTCTCCGAGC	TCTTCTTCTT	CCACCTCCTG	GCTGGGGCAG	ACTGCTTCTT	GCTGACCATC	ATGGCCTATG	160							
161	ACCGCTATCT	GGCCATCTGC	CAGTCCCTCA	CCTACAGCAG	CCGCATGAGC	TGGGGAATCC	AGCAAAGCCCT	GGTGGGCATG	240							
241	TCATGTGTCT	TTTCCTTCAC	CAATGCAC TG	ACCCAAAC TG	TTGCCCTGTC	TCCTCTTAAC	TTCTGTGGCC	CCAATGTGAT	320							
321	CAATCACTTC	TACTGTGACC	TCCCACAGCC	CTTCCAGCTC	TCCTGGGCCA	GTGTTCACTC	CAATGGGCAG	TTGCTGTTTG	400							
401	TAGCAGCAGC	TTTCATGGGT	GTGGCCCCCT	TGGTCCTCAT	CAC TGTGTC	TATGCCCATG	TGGCAGCTGC	AGTCTCGGA	480							
481	ATCCGCTCTG	CAGAGGGCAA	AAAGAAAGCC	TTCTCCACGT	GTAGTTCCCA	CCTCACTGTG	GTGGGCATCT	TCTATGGGAC	560							
561	GGCGTCTTC	AGCTACACAA	GGCTGGGTT C	AGTGGAGTCT	TCGGACAAGG	ACAAAGGCAT	TGGCATCCTC	AACACTGTCA	640							
641	TCAGCCCC								648							

**FIG. 4a**

17-219

[illegible]

**FIG. 4b**

9/16

17-40

1 GTGCTGGATG TTGGGTGCAT CAGCGTCACT GTTCCATCAA | 40  
81 TGGGGCCTGC CTTACCCAGC TCTTCTTCTT CCATCTGTTT | 50  
161 ACCAATTCTT GGGCATCTGC CCGCCCTCTCA CCTACAGCAC | 60  
241 TCCTGGGCTT GTGCTTTTAC CAAGCACTG ACCACACTG | 70  
321 CAATCACTTC TACTGTGACC TCCACAGCT CTTCAGTTT | 80  
401 CTGTGGGTTT TATAATGGCA GGTACCCCTCA TGGCTCTCAT | 90  
481 ATCCGCTCTG TAGAGGGCAG GAAGAAAGCC TTCTCCACAT | 100  
561 AGGTATCTTT AACTATATGC GACTGGGTTT AACCAAGCTT | 110  
641 TCAATCCC | 120

FIG. 4c

17-82

1 GTGCTGGATG TTGGGTGCAT CAGCGTCACT GTTCCATCAA | 40  
81 TGGGGCCTGC CTTACCCAGC TCTTCTTCTT CCATCTGTTT | 50  
161 ACCAATTCTT GGGCATCTGC CCGTCCTCTCA CCTACAGCAC | 60  
241 TCCTGGGCTT GTGCTTTTAC CAAGCACTG ACCACACTG | 70  
321 CAATCACTTC TACTGTGACC TCCACAGCT CTTCAGTTT | 80  
401 CTGTGGGTTT TATAATGGCA GGTACCCCTCA TGGCTCTCAT | 90  
481 ATTCGCTCTG TAGAGGGCAG GAAGAAAGCC TTCTCCACAT | 100  
561 AGGTATCTTC AACTACATGA GACTGGGTTT AGAGGAGGCT | 110  
641 TCAACCTT | 120

FIG. 4d

10/16

17-228

```

1  GTGCTGGATG TCGGATGTAT CACTGTCACT GTTCTTGCAA | 40
81  TGACGCCTGC CTCTCCAGC TCTTCTTCTT CCACCTTCTG | 50
161 ACCGACTCCT GGGCATCTGC CAGCCCTCA CCTACAGGAC | 60
241 TCCTTGGCTT GTGCTTTCAC CAAGCACTG ACCACACTG | 70
321 CAATCACTTC TACTGTGACC TCCCACAGCT CTTCCAGCTC | 80
401 CTGTGGGTTT CATCATGGCA GGCACACCTT TGGTTCTCAT |
481 ATCCGTTTCAG TGGAGGGCCG AAAGAAGGCC TTCTCCACGT |
561 AGGTATCTTC AACTACATGA GACTGGGTTT AGAGGAGGCT |
641 TCAACCTT

```

FIG. 4e

17-207

```

1  GTGCTGGATG TCGGATGTAT CACTGTCACT GTTCTTGCAA | 40
81  TGACGCCTGC CTCTCCAGC TCTTCTTCTT CCACCTTCTG | 50
161 ACCGACTCCT GGGCATCTGT CAGGCCCTCA CCTACAGGAC | 60
241 TCCTTGGGCTT GTGCTTTCAC CAAGCACTG ACCACACTG | 70
321 CAATCACTTC TACTGTGACC TCCCACAGCT CTTCAAGCTC | 80
401 CTGTGGGTTT AATAATGGCA GGTACCCCTT TGGCTCTCAT |
481 ATTCGCTCTG TAGAGGGCAG GAAGAAGCC TTCTCCACAT |
561 AGGTATCTTC AACTATATGC GACTGGGTTT AACCAAGCTT |
641 TCAATCCC

```

FIG. 4f



11 / 16

17-201

```

1  GTGCTGGATG TCGGATGTAT CACTGTCACT GTTCCTGCAA      | 40
81  TGACGCGTGC CTCTCCAGC TCCTCTTCTT CCACCTTCTG      | 50
161 ACCGACTCCT GGCATCTGC CAGCCCTCA CCTACAGcAC      | 60
241 TCCTGGGCTT GTGCCTTCAC CAAGCACTG ACCACACTG      | 70
321 CAATCACTTC TACTGTGAAC TCCCACAGGT CTTCAGCTC      | 80
401 TAGCAGCAGC CTTTCATGGCT GTGGCACCTT TGGTCTTCAT      |
481 ATCCGCTCTG CTGAGGGCAG AAAGAAGGCC TTCTCCACAT      |
561 AGGTGTCTTC AGCTACATGA GGCTGGGTTT AGTGAATCT      |
641 TCAACCCC      | 10      | 20      | 30      | 40      | 50      | 60      | 70      | 80

```

FIG. 4g

17-22

```

1  TTCACTGACC TCTTCTTTGT CACCAACACA ATCCCCAAGA      | 40
81  CACAGGGTGT CTGACACAGC TCTACTTCTT GGTCTCCTTG      | 50
161 ATCGCTATGT GGCCATCTGC CGTCCCTCC ACTATGTAC      | 60
241 TGTGGGTGT TCTTGCCCT CTATGGCTC ATCCATATCC      | 70
321 CCACTACCTC TTCTGTGAGA TGTACTTCTT GCTAAGGCTG      | 80
401 CCACGGGCTG CTTTCATCTTC CTCATCCCTT TAGGTTTCAT      |
481 ATACCCTCAG CCACTGGGAA GTACAAAGCC TTCTCCACCT      |
561 TCTGGGTATG GTGTACCTGC AGCCCTTCCA AACCTACTCC      |
641 CGCC      | 10      | 20      | 30      | 40      | 50      | 60      | 70      | 80

```

FIG. 4h

12/16

17-90

```

1  TTCACGTGACC  TCTTCTTTGT  CACCAACACA  ATCCCCAAGA  |  40
81  CACAGGGTGT  CTGACACAGC  TCTACTTCTT  GGTCTCCTTG  |  80
161 ATCGCTATGT  GGCCATCTGC  CGTCCCTCTC  ACTATGTCAC  |  120
241 TGTGGGTGT  TCTCTGCCCT  CTATGGCCTC  ATCCATATCC  |  160
321 CCACTACCTC  TCTGTGAGA  TGTACTTCTT  GCTAAGGCTG  |  200
401 CCACGGGCTG  CTTTATCTTC  CTCATCCCTT  TAGTTTCAT   |  240
481 ATACCCTCAG  CCACTGGGAA  GTACAAAGCC  TTCTCCACCT  |  280
561 GGGTATGGTG  TACCTGCAGC  CCCTCCAAAC  CTACTCCATG  |  320
641 C

```

FIG. 4i

17-4

SUBSTITUTE SHEET (RULE 26)

```

1  TTCACGTGACC  TCTTCTTTGT  CACCAACACA  ATCCCCAAGA  |  40
81  TGCAGGGTGT  CTGACACAGC  TCTACTTCTT  GGTCTCCTTG  |  80
161 ACCGCTATGT  GGCCATCTGC  TGCCTCTCTC  ACTACACACC  |  120
241 TGTGGGTCC  TATCCGTCCT  CTATGGCCTC  ATACACACCC  |  160
321 CCACTACATC  TTCTGTGAGA  TGTATGTATT  GCTGAGGATG  |  200
401 CCACAGGCTG  CTTTATCTTC  CTCATTCCTT  TTGGATTCTG  |  240
481 ATACCCTCAG  TCTTAAGAA  ATACAAAGCC  TTCTCCACCT  |  280
561 ACTTTGTATG  GTATACCTAA  AGCCCTCTCA  TACCTACTCT  |  320
641 CACCC

```

FIG. 4j

		10		20		30		40		50		60		70		80
1	TTCACTGACC	TCCTCTTTGT	CACCAACACA	ATCCCAAGA	TGCTGGTGAA	CTTCCAGTCC	CAGAACAAAG	CCATCTCCTA	80							
81	TGCAGGGTGT	CTGACACAGC	TCTACTTCT	GGTCTCCTTG	GTGACCCCTGG	ACAACCTCAT	CCTGGCCGTG	ATGGCGTATG	160							
161	ATCGCTATGT	GGCCATCTGC	TGCCCCCTCC	ACTATGTAC	AGCCATGAGC	CCTGGGCTCT	GTGTCTTGCT	CCTCTCCTTG	240							
241	TGTTGGGGGC	TGCTGTGTTCT	CTATGGCCTC	CTCTCACCT	TCCTCCTGAC	CAGGGTGACC	TTCTGTGGGC	CTCGAGAGAT	320							
321	CCACTACCTC	TTCTGTGACA	TGTACATCCT	GCTGTGGCTG	GCATGTTCCA	ACACCCACAT	CATTACACACA	GCATTGATTG	400							
401	CCACTGGCTG	CTTCATCTTC	CTCACCCCTCT	TAGGGTTTCAT	GACCACATCC	TATGTACGTA	TTGTCAGAAC	CATCCTTCAA	480							
481	ATGCCCTCGG	CCTCTAAGAA	ATACAAAAC	TTCTCTACCT	GTGCCCTCGCA	TTTGGGTGTG	GTCTCCCTCT	TTTATGGGAC	560							
561	GCTTGCTATG	GTGTACCTGC	AGCCCCCTCCA	TACCTACTCC	ATGAAGGACT	CAGTAGCCAC	AGTGATGTAT	GCTGTGCTGA	640							
641	CACCT								645							

**FIG. 4k**

		10		20		30		40		50		60		70		80
1	TCTCTGACC	TCTGCTTCTC	TTCCGTGACC	ATTCCAAGT	TGTTACAGAA	CATGCAGAAC	CAGGACCCAT	CCATCCCCCA	80							
81	TGGCGACTGC	CTGACCCAAA	TGTACTTCCT	CCITGTTATT	GGAGACCTGG	AGAGCTTCCT	CCTTGCGCC	ATGGCCTATG	160							
161	ACCGCTATGT	GGCCATCTGC	TTCCCCTGCG	ACTACACCGC	CATCATGAGC	CCCATGCTCT	GTCTCGCCCT	GTTGGCGCTG	240							
241	TCCTGGGTGC	TGACCACCTT	CCATGCCATG	TTACACACTT	TACTCATGGC	CAGGTTGTGT	TTTTGTGCAG	ACAATGTGAT	320							
321	CCCCCACCTT	TTCTGTGATA	TGCTGTGCTCT	GCTGAAGCTG	GCCTTCTCTG	ACACTCGAGT	TAATGAATGG	GTGATATTTA	400							
401	TCATGGGAGG	GCTCATTCCT	GTCATCCCAT	TCCTACTCAT	CCTTGGGTCC	TATGCAAGAA	TTGCTCTCTC	CATCCTCAAG	480							
481	GTCCCTTCCT	CTAAGGGTAT	CTGCAAGGCC	TTCTCTACTT	GTGGCTCCCA	CCTGTCTGTG	GTGTCACTGT	TCTATGGAAC	560							
561	CGTTATTGGT	CTCTACTTAT	GCTCATCAGC	TAATAGTTCT	ACTCTAAAGG	ACACTGTGAT	GGCTATGATG	TACACTGTGG	640							
641	TGACCCCC								648							

FIG. 4L

14/16

17-32

1	TTCTCTGACC	10	TCTGCTTCTC	20	TTCCGTGACC	30	ATTCCCAAGT	40	TGTTACAGAA	50	CATGCAGAAC	60	CAGGACCCAT	70	CCATCCCCCTA	80
81	TGCGGACTGC		CTGACCCCAA		TGTACTTCTT		CCTGTTATTT		GGAGACCTGG		AGAGCTTCTT		CCTTGTGGCC		ATGGCCTATG	160
161	ACCGCTATGT		GGCCATCTGC		TTCCCCCTGC		ACTACACCGC		CATCATGAGC		CCCATGCTCT		GTCTCACCCCT		GGTGGCGCTG	240
241	TCCTGGGTGC		TGACCACCTT		CCATGCCATG		TTACACACTT		TACTCATGGC		CAGGTTGTGT		TTTTGTGCAG		ACAATGTGAT	320
321	CCCCCACTTT		TTCTGTGATA		TGCTGCTCT		GCTGAAGCTG		GCCTTCTCTG		ACACTCGAGT		TAATGAATGG		GTGATATTTA	400
401	TCATGGGAGG		GGTCATTCTT		GTCATCCCAT		TCCTACTCAT		CCTTGGGTCC		TATGCAAGAA		TTGTCTCCTC		CATCCTCAAG	480
481	GTCCCTTCTT		CTAAGGGTAT		CTGCAAGGCC		TTCTCTACTT		GTGGCTCCCA		CCTGTCTGTG		GTGTCACTGT		TCTATGGAAC	560
561	CGTTATTGGT		CTCTACTTAT		GCTCATCAGC		TAATAGTTCT		ACTCTAAAGG		ACACTGTCTAT		GGCTATGATG		TACACTGTGG	640
641	TGACCCCCC															648

FIG. 4m

17-93

1	TTCTCTGACC	10	TCTGCTTTTC	20	CTCAGTCACA	30	ATGCCCAAAT	40	TGCTGCAGAA	50	CATGCAGAAC	60	CAAGACCCAT	70	CCATCCCCCTA	80
81	TGCAGACTGT		CTGACCCCAA		TGTACTTCTT		CTTGTAATTT		TCGGATCTAG		AGAGCTTCTT		CCTTGTGGCC		ATGGCCTATG	160
161	ACCGCTATGT		GGCCATCTGC		TTCCCCATGC		ACTACACCGC		CATCTGCTTC		CTCCTGCACT		ACACCGCCAT		CATGAGCCCC	240
241	ATGCTCTGTC		TCTCCGTGGT		GGCGCTGTCC		TGGGTGCTGA		CCACCTTCCA		TGCCATGTTA		CACACTTTAC		TCATGGCCAG	320
321	GTGTGTTTTT		TGTGCAGACA		ATGTGATCCC		CCACTTTTTC		TGTGATATGT		CTGCTCTGCT		GAAGCTGGCC		TGCTCTGACA	400
401	CTCGAGTTAA		TGAATGGGTG		ATATTATCA		TGGGAGGGCT		CATTCTTGTC		ATCCCATTC		TACTCATCCT		TGGGTCTTAT	480
481	GCAAGAAATG		TCTCCTCCAT		CCTCAAGGTC		CCTTCTTCTA		AGGGTATCTG		CAAGGCCCTTC		TCTACTTTGTG		GCTCCACCT	560
561	CTCTGTGGTG		TCACTGTTCT		ATGGGACCGT		TATTGGTCTC		TACTTATGCC		CATCAGCTAA		TAGTTTCTACT		CTAAAGGACA	640
641	CTGTCTATGC		TATGATGTAC		ACTGTGGTGA		CCCTT									675

FIG. 4n

15/16

17-210

1	TTCTCTGACC	10	1	20	1	30	1	40	1	50	1	60	1	70	1	80
81	TGCGGACTGC	CTGGCTCAGA	TGACTTTTCA	TCTGTTTTAT	ATGCCCAAAT	CTCGGTACACA	CTGGGTACACA	ATGCCCAAAT	TGCTGCAGAA	CATGCAGAGC	CAAAACCCAT	CCATCCCCCTT	80			
161	ACTGCTATGT	GGCTATTTGC	TTTCCTCTGC	ACTACACCAC	TTGCAACCTT	CCATGACCTG	CCATGACCTG	TTGCAACCTT	TATCATGAGC	CCCAAAGTGT	GGCTTGCTCT	GCTGACACTC	240			
241	TCCTGGCTGT	TGACCACTGC	CCATGACCTG	TTGCAACCTT	CTGCAAGCTG	CATCTACCTT	CTGCAAGCTG	CTGCAAGCTG	TGCTTATGGG	CAGCCTGTCC	TTTTGTGCTG	AGAATGTGAT	320			
321	TCCTCACTTT	TTCTGTGATA	CATCTACCTT	CTGCAAGCTG	CTGCAAGCTG	CTGCAAGCTG	CTGCAAGCTG	CTGCAAGCTG	GCCTGTCTCA	ACACGCAAGT	CAATGGGTGG	GTGATGTTTT	400			
401	TCATGGGCGG	GCTCATCCTT	GTCATCCCAT	TCTTACTCCT	TCTTACTCCT	GTCATCCCAT	GTCATCCCAT	TCTTACTCCT	CATCATGTCC	TGTGCAAGAA	TGCTCTCCAC	CATCCTCAGG	480			
481	GTCCCTTCCA	CTGGGGGCAT	CCAGAAGGCT	TTCTCCACCT	TTCTCCACCT	CCAGAAGGCT	CCAGAAGGCT	TTCTCCACCT	GTGGCCCCCA	CCTCTCTGTG	GTGTCTCTCT	TCTATGGGAC	560			
561	AATTATTGGT	CTCTACTTGT	GCCCATTTGAC	GAATCATAAC	GAATCATAAC	GCCCATTTGAC	GCCCATTTGAC	GAATCATAAC	ACTGTGAAGG	ACACTGTCTAT	GGCTGTGATG	TACACTGGGG	640			
641	TGACCCAC												648			

FIG. 40

17-209

1	CTTGACAGATG	10	1	20	1	30	1	40	1	50	1	60	1	70	1	80
81	CTTAGGGTGT	CTACTACAGT	TGATTTTTTT	CATGTTATTT	GTCCCTAAGA	GTCCCTAAGA	GTCCCTAAGA	GTCCCTAAGA	TGCTGGCAAA	CATACAGATC	CAGAGTCAGG	CCATCTCCTA	80			
161	ACTGCTACGT	GGGCATATGC	CACCCACTTC	ATTACATTCT	ATTACATTCT	CACCCACTTC	CACCCACTTC	ATTACATTCT	GATCATGAGC	CCTGGGCTCT	GCATCTTCTT	CGTGTCTGCA	160			
241	TCCTGGATCA	TGAATGCCCT	CCACTCCCTT	CTACACACAC	CTACACACAC	CCACTCCCTT	CCACTCCCTT	CTACACACAC	TTCTGATGAA	CAGCCTGTCC	TTCTGGCGAA	ACCATGAGAT	240			
321	CCACACCTTC	TTCTGTGACA	TCAATCCCTT	CCTGAGTCTG	CCTGAGTCTG	TCAATCCCTT	TCAATCCCTT	CCTGAGTCTG	TCCTGCACAG	ACCCCTTCAC	CAATGAGCTG	GTGATCTTCA	320			
401	TCACTGGGGG	TCTCACAGGA	CTCATTTGTG	TGCTTTGCCT	TGCTTTGCCT	CTCATTTGTG	CTCATTTGTG	TGCTTTGCCT	GATTATCTCT	TACACGAACG	TTTTCTCGAC	CATCCTGAAG	400			
481	ATCCCATCAG	CTCAGGGGAA	GCGGAAAGCC	TTTTCCACCT	TTTTCCACCT	CTCAGGGGAA	CTCAGGGGAA	TTTTCCACCT	GCAGCTCTCA	TCTCTCCGTG	GTCTCTCTCT	TCTTTGGGAC	480			
561	TTCTTTTTGT	GTTGATTTCA	GTTCTCCCTC	AACCACTCG	AACCACTCG	GTTGATTTCA	GTTGATTTCA	AACCACTCG	GCCGAGAAAG	ACACAGTTGC	ATCAGTGATG	TACACAGTGG	560			
641	TAACTCCA												648			

FIG. 4p

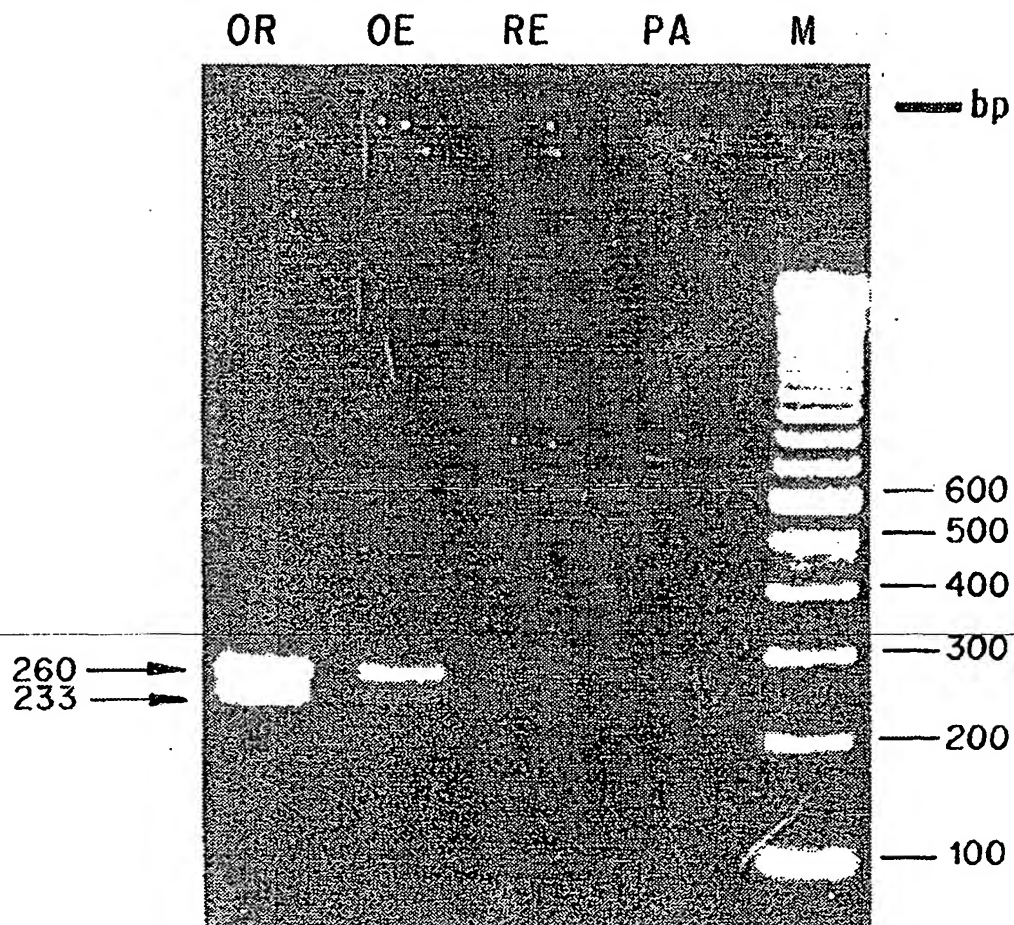


FIG. 5

## INTERNATIONAL SEARCH REPORT

International application No.

PCT/US94/14554

## A. CLASSIFICATION OF SUBJECT MATTER

IPC(6) : Please See Extra Sheet.

US CL : Please See Extra Sheet.

According to International Patent Classification (IPC) or to both national classification and IPC

## B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

U.S. : 435/69.1, 70.1, 70.3, 71.1, 240.1, 240.2, 320.1; 530/300, 350; 536/23.1, 23.5, 24.31

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

Please See Extra Sheet.

## C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	Nature, Volume 355, issued 30 January 1992, M. Parmentier et al, "Expression of Members of the Putative Olfactory Receptor Gene Family in Mammalian Germ Cells", pages 453-45, see entire documents, especially page 454.	1-8
Y	Nature, Volume 361, issued 28 January 1993; K. Raming et al, "Cloning and Expression of Odorant Receptors", pages 353-356, see entire document, especially page 354.	1-8
Y	Cell, Volume 65, issued 05 April 1991, L. Buck et al, "A Novel Multigene Family May Encode Odorant Receptors: A Molecular Basis for Odor Recognition", pages 175-187, see entire document, especially pages 179-180.	1-8



Further documents are listed in the continuation of Box C.



See patent family annex.

Special categories of cited documents:		T	later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
*A*	document defining the general state of the art which is not considered to be of particular relevance	X	document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
*E*	earlier document published on or after the international filing date	Y	document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
*L*	document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)	A*	document member of the same patent family
*O*	document referring to an oral disclosure, use, exhibition or other means		
*P*	document published prior to the international filing date but later than the priority date claimed		

Date of the actual completion of the international search

19 MARCH 1995

Date of mailing of the international search report

03 APR 1995

Name and mailing address of the ISA/US  
Commissioner of Patents and Trademarks  
Box PCT  
Washington, D.C. 20231

Facsimile No. (703) 305-3230

Authorized officer

ARIE M. MICHELSON, PH.D.

Telephone No. (703) 308-0196

International application No.  
PCT/US94/14554

International application No.  
PCT/US94/14554

[illegible]



# INTERNATIONAL SEARCH REPORT

International application No.  
PCT/US94/14554

## Box I Observations where certain claims were found unsearchable (Continuation of item 1 of first sheet)

This international report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claims Nos.:  
because they relate to subject matter not required to be searched by this Authority, namely:
  
2. ☐ Claims Nos.:  
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:
  
3. ☐ Claims Nos.:  
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

## Box II Observations where unity of invention is lacking (Continuation of item 2 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

Please see Extra Sheet

1. ☐ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.
2. ☐ As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.
3. ☐ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:
  
4. ☒ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:  
1-8

Remark on Protest

- ☐ The additional search fees were accompanied by the applicant's protest.  
☐ No protest accompanied the payment of additional search fees.

## BOX II OBSERVATIONS WHERE UNITY OF INVENTION IS LACKING:

- I. Claims 1-8, directed to nucleotide sequences and corresponding protein sequences, host cells and recombinant method of producing the corresponding protein sequences, classified in U.S. Class 435, subclass 69.1 and Class 530, subclass 350, and others.
- II. Claim 9, directed to an antibody, classified in U.S. Class 530, subclass 387.9.

Groups I and II are materially distinct compositions of matter that are distinguished, each from the other, by their special technical features: the polynucleotides and corresponding polypeptides, host cells and method of producing the corresponding proteins of Group I, and the antibody of Group II, have materially different structures and functions. Accordingly, the claims are not so linked by a special technical feature within the meaning of PCT Rule 13.2 so as to form a single inventive concept.

## A. CLASSIFICATION OF SUBJECT MATTER:

IPC (6):

C07H 21/00, 21/04; C07K 2/00, 14/00; C12N 5/10, 5/16, 7/01, 15/00, 15/10, 15/11, 15/12, 15/63; C12P 1/00, 21/00, 21/02

## A. CLASSIFICATION OF SUBJECT MATTER:

US CL :

435/69.1, 70.1, 70.3, 71.1, 240.1, 240.2, 320.1; 530/300, 350; 536/23.1, 23.5, 24.31

## B. FIELDS SEARCHED

Electronic data bases consulted (Name of data base and where practicable terms used):

APS; STN files Biosis, Medline, EMBASE, CA, WPIDS; sequence search through Intelligenetics; databanks EMBL-NEW, GenBank 86 & NEW, UEMBL. Search terms for STN search: (olfactory(w)receptor#) and clon?, cDNA, DNA, amino acid sequence#, genomic DNA, recombinant DNA, nucleic acid#.